“Tensor Regression with Applications in Neuroimaging Data Analysis”

Hua Zhou, Lexin Li, & Hongtu Zhu

Jennifer Kirk

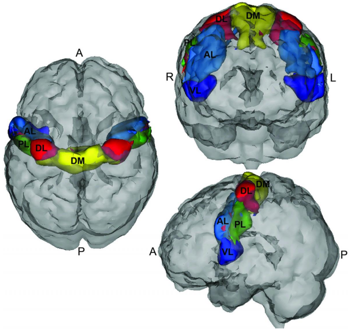
May 30, 2013

Outline

- Motivation
- Tensor Regression
- Simulation Results
- Discussion

Scientific Motivation

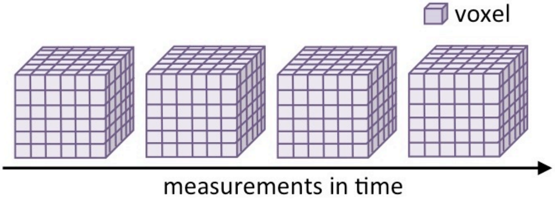
- Mental health disorders are difficult to diagnose and treat
- Physiology of the brain is not well understood
- Neuroimaging can elucidate the brain's physiology
- Several types of neuroimaging, e.g. PET, MRI, **fMRI**



Brain Areas Associated with ADHD from fMRI
Image Source: MIT Tech Review

Statistical Motivation

- fMRI data: 4-D array (tensor) with spatial and temporal correlation
- Naive approach: use image as vector covariate
  - Lots of data $\Longrightarrow$ **lots of parameters** ($\approx$ 16 million)
  - Ignores **spatial and temporal correlation**
- **New Method**: Extend GLM to use **fMRI image as one covariate observation** in regression model



One fMRI Observation from One Subject

Current Methods

- Voxel Based Methods
  - Analysis of each voxel as response variable
  - Assumes voxels independent–ignores spatial correlation
- Functional Data Methods
  - Collapses data into one parameter function
  - Commonly used for 2-D data, extension to 3-D data is complex
- Two-Stage Reduction Methods
  - Reduce the dimension of the data, possibly more than once, then model the reduced data
  - Theoretical properties are intractable and reduction maybe unrelated to response

Special Case: Matrix Covariates

Recall:

- Outcome $Y_i \sim$ univariate exponential family
- Vector covariate: $\mathbf{z}_i$
- $\mathbf{X_i}$ is a $p \times q$ matrix
- β_1^{T} is a $1 \times p$ vector
- β_2 is a $q \times 1$ vector

$$
\begin{aligned}
g(\mu_i) &= \alpha + \gamma^{\mathsf{T}}\mathbf{Z_i} + \beta_1^{\mathsf{T}}\mathbf{X_i}\beta_2 \\
&= \alpha + \gamma^{\mathsf{T}}\mathbf{Z_i} + \langle(\beta_2 \odot \beta_1), \mathrm{vec}(\mathbf{X_i})\rangle
\end{aligned}
$$

where $(\beta_2 \odot \beta_1)$ is a $pq \times 1$ vector, $\langle\cdot\rangle$ is the inner product, and $\mathrm{vec}(\mathbf{X_i})$ is the vector form of $\mathbf{X_i}$

Tensor Notation

- Order: the number of indices need to describe the tensor
- Kronecker Product: A is $m \times p$, B is $n \times q$:

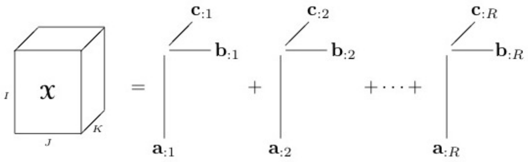
$$\mathbf{A} \otimes \mathbf{B}_{mn \times pq} \equiv \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,p}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,p}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}\mathbf{B} & a_{m,2}\mathbf{B} & \cdots & a_{m,p}\mathbf{B} \end{bmatrix}$$

- Khatri-Rao Product: A is $m \times p$, B is $n \times p$:

$$\mathbf{A} \odot \mathbf{B}_{mn \times p} \equiv [\mathbf{a}_{\cdot 1} \otimes \mathbf{b}_{\cdot 1} \cdots \mathbf{a}_{\cdot p} \otimes \mathbf{b}_{\cdot p}]$$

Rank-R Decomposition

- If $\mathbf{X}$ is an $I \times J \times K$ (order 3) tensor and $\mathbf{A}_{I \times R}$, $\mathbf{B}_{J \times R}$, $\mathbf{C}_{K \times R}$ are matrices then $\mathbf{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ means

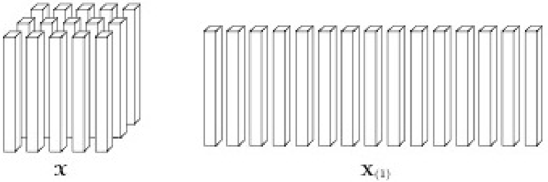


- If $\mathbf{X}$ is an $I_1 \times \ldots \times I_D$ (order D) tensor, then the rank-R decomposition is

$$\mathbf{X} = [\![\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_D]\!] = \sum_{r=1}^{R} \mathbf{a}_1^{(r)} \circ \ldots \circ \mathbf{a}_D^{(r)}$$

Mode-d Matricization

- Denoted $\mathbf{X}_{(d)}$
- $\mathbf{X}$ is an $I \times J \times K$ (order 3) tensor then $\mathbf{X}_{(1)}$:



- In general, we "spread out" the tensor, keeping the d^{th} dimension, to get a matrix

Rank-R Generalized Linear Tensor Regression

- Outcome: $Y_i \sim$ univariate exponential family
- Vector covariate: $\mathbf{z_i}$
- Tensor covariate: $\mathbf{X_i}$ (Order D: $I_1 \times \ldots \times I_D$)
- Assume tensor $\mathbf{B}$ has a rank-R decomposition

$$[\![\mathbf{B}_1, \ldots, \mathbf{B}_D]\!]$$

  where $\mathbf{B}_d$ is $I_d \times R$ matrix
- Link function:

$$g(\mu_i) = \alpha + \gamma^{\mathsf{T}} \mathbf{z}_i + \langle (\mathbf{B}_D \odot \ldots \odot \mathbf{B}_1) \mathbf{1}_R, \mathrm{vec}(\mathbf{X_i}) \rangle$$

Parameter Estimation

- Maximum Likelihood Estimation
- Estimation Algorithm
  1. Set $\mathbf{B}^{(0)} = 0$ & estimate $\hat{\alpha}^{(0)}$, $\hat{\gamma}^{(0)}$
  2. Set $\alpha = \hat{\alpha}^{(n-1)}$, $\gamma = \hat{\gamma}^{(n-1)}$ & for each $\mathbf{B}_d$:
     - Set $\mathbf{B}_k = \hat{\mathbf{B}}_k^{(n)}$, $k < d$
     - Set $\mathbf{B}_k = \hat{\mathbf{B}}_k^{(n-1)}$, $k > d$
     - Estimate $\hat{\mathbf{B}}_d$
  3. Estimate $\hat{\alpha}^{(n)}$ and $\hat{\gamma}^{(n)}$, assuming $\mathbf{B}_d = \hat{\mathbf{B}}_d^{(n)}$ for all d
  4. Iterate 2–3 until the likelihood converges

$$g(\mu_i) = \alpha + \gamma^{\mathsf{T}} \mathbf{z}_i + \langle \mathbf{B}_d, \mathbf{X}_{\mathbf{i}(d)} (\mathbf{B}_D \odot \ldots \odot \mathbf{B}_{d+1} \odot \mathbf{B}_{d-1} \odot \ldots \odot \mathbf{B}_1) \rangle$$

Simulation: Set up

- 100 replications
- 1000 observations
- $\mathbf{X_i} \sim N_{20\times20}(\mathbf{0}, \mathbf{I}, \mathbf{I})$
- $\mathbf{B}$: Image Parameter
- $\mu_i = \langle \mathbf{B}, \mathbf{X_i} \rangle$

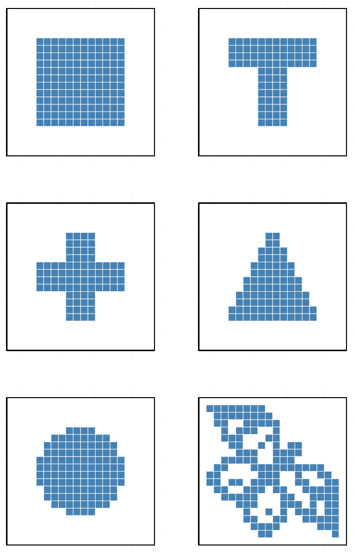
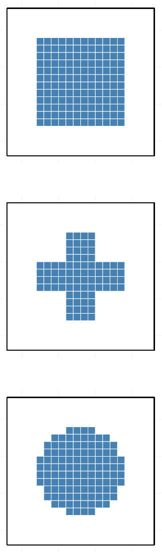
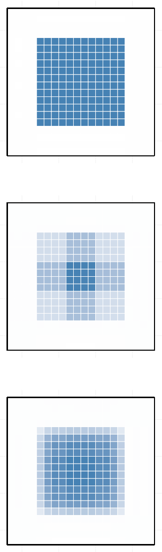


Image Parameters

Simulation: Unbiased

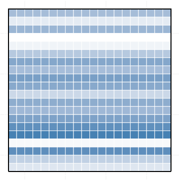


True Parameters

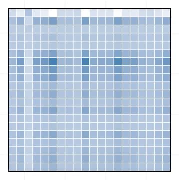


Rank-1 Models

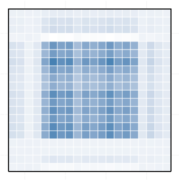
Simulation: Unbiased



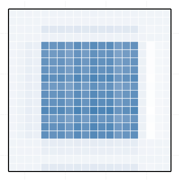
N = 10



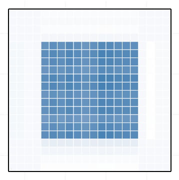
N = 25



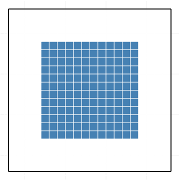
N = 40



N = 50

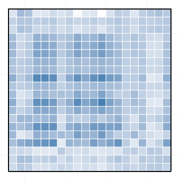


N = 100

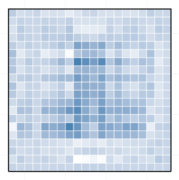


Truth

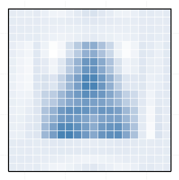
Simulation: Unbiased



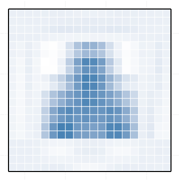
N = 50



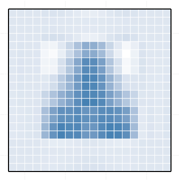
N = 100



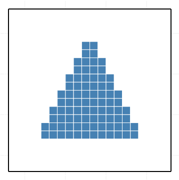
N = 500



N = 1000



N = 2000



Truth

Score and Information

- Score

$$\nabla \ell(\mathbf{B}_1, \ldots, \mathbf{B}_d) = \frac{(y-\mu)}{\sigma^2} \underbrace{\mu'(\eta)[\mathbf{J}_1, \ldots, \mathbf{J}_D]^\mathsf{T} vec(\mathbf{X})}_{\frac{d\mu}{d\eta}\frac{d\eta}{d\beta}}$$

- Information

$$\mathbf{I}(\mathbf{B}_1, \ldots, \mathbf{B}_D) = \frac{[\mu'(\eta)]^2}{\sigma^2}[\mathbf{J}_1, \ldots, \mathbf{J}_D]^\mathsf{T}(vec\mathbf{X})(vec\mathbf{X})^\mathsf{T}[\mathbf{J}_1, \ldots, \mathbf{J}_D]$$

Asymptotic Normality

For an interior point, $\mathbf{B}_0 = [\![\mathbf{B}_{01}, \ldots, \mathbf{B}_{0D}]\!]$

$$\sqrt{n}[vec(\hat{\mathbf{B}}_{n1}, \ldots, \hat{\mathbf{B}}_{nD}) - vec(\mathbf{B}_{01}, \ldots, \mathbf{B}_{0D})]$$

converges to

$$N(\mathbf{0}, \mathbf{I}^{-1}(\mathbf{B}_{01}, \ldots, \mathbf{B}_{0D}))$$

Non-Identifiability

Two types of indeterminacy:

- Scaling & permutation indeterminacy
- Non-unique Rank-R decomposition

Discussion

- Tensor parameter decomposition may not be interpretable
- Asymptotics require large sample size ($n > p$)
- Computation speed

Summary

- Analysis of complex neuroimages is important for understanding brain physiology
- fMRI data is complex: 4-D array with spatial and temporal correlation
- Current analysis methods ignore one or more of these features
- Tensor regression extends GLM to array covariates
- Extend GLM framework to tensor covariates
- Classical theory results hold, but large sample size required